Natural Language Processing, Pragmatics, and Verbal Behavior

Chris Cherpas
Fremont, CA

Natural Language Processing (NLP) is that part of Artificial Intelligence (AI) concerned with endowing computers with verbal and listener repertoires, so that people can interact with them more easily. Most attention has been given to accurately parsing and generating syntactic structures, although NLP researchers are finding ways of handling the semantic content of language as well. It is increasingly apparent that understanding the pragmatic (contextual and consequential) dimension of natural language is critical for producing effective NLP systems. While there are some techniques for applying pragmatics in computer systems, they are piecemeal, crude, and lack an integrated theoretical foundation. Unfortunately, there is little awareness that Skinner's (1957) *Verbal Behavior* provides an extensive, principled pragmatic analysis of language. The implications of Skinner's functional analysis for NLP and for verbal aspects of epistemology lead to a proposal for a "user expert" — a computer system whose area of expertise is the long-term computer user. The evolutionary nature of behavior suggests an AI technology known as genetic algorithms/programming for implementing such a system.

Natural Language Processing (NLP) utilizes the concepts and techniques of Artificial Intelligence (AI), philosophy, linguistics, psychology, and related disciplines to create computer systems which mimic human language capabilities. Some of the practical applications of NLP are question-answering and database retrieval, text analysis and generation, and machine translation (Winograd, 1983, pp. 359-360).

The "N" in NLP denotes a special concern for *natural* language (as in conversations between people), in contrast to *artificial* languages which are designed by people to program and control the operations of computers. The typical task for an aspiring NLP system is to somehow interpret peoples' natural language inputs to a computer without their having to learn an artificial language. A complementary linguistic talent of an NLP system is to make the computer generate text or simulate speech which can be read or heard as if it were written or spoken by another person.

Reprints may be obtained from the author, Computer Curriculum Corporation, 1287 Lawrence Station Road, Sunnyvale, CA 94089.

Components of NLP

For historical and practical reasons, NLP systems are divided into at least four different functional components:

1. morphological/lexical: providing the basic language elements or vocabulary, such as words, their roots, and inflections;
2. syntactic: for grouping and sequencing elements within samples of language (usually sentences);
3. semantic: for knowing the meaning of an utterance, usually defined in terms of the "truth value" of the logical propositions that are thought to be expressed by sentences;
4. pragmatic: for understanding the context and pupose of an utterance.

By no means does every NLP system attempt to implement all four of these components. In fact, there is no consensus on the extent to which these functions should even be separated, or whether some functions are primary while others are secondary. For now, we will consider the separate challenges facing each of the components for NLP. As we progress from the lexical through the syntactic and the semantic to the pragmatic, we will find that each component leaves a gap or other problem which the next component partially resolves.

The earliest attempts at machine translation may have aspired to little more than exercising the first of the four components, that is, looking up each word of one language and replacing it with a corresponding dictionary entry in another language. Even taking advantage of known word roots and inflections, this approach could have only limited success. How a word fits into a given clause or sentence is simply beyond the analysis of individual words. Consider how the word "fast" is interpreted differently in each of the following sentences:

"His conversion to the monastic life was awfully fast."
"His fast will endure for three days."
"He can fast longer than anyone I know."

It is the second component of an NLP system, syntax, which contributes the ability to identify the classes to which words belong in a sentence (e.g., adjective, noun, verb), and constrains the possible groupings and orderings of words and phrases to those allowable in a given language. Syntactic analysis and generation are by far the most thoroughly researched areas in NLP and have been implemented in a variety of ways. Probably the most well known approach to generating sentences is the use of "context-free" and "transformational" grammars (Chomsky, 1957, 1965), while a typical approach for analyzing sentences is the use of "augmented transition networks" (e.g., Woods, 1970, 1973). When a sentence is parsed (analyzed), all of its words are categorized and organized into a "phrase structure."

Unfortunately, ambiguities arise since more than one phrase structure can be interpreted from a given sentence. "Flying planes can be dangerous" might be uttered by a pilot who has just narrowly escaped crashing into a mountain — or by a person standing on that mountain. Syntactic analysis is further limited in that it is usually confined to individual, complete sentences — not necessarily the stuff of which conversations are made. Even with a more flexible approach to syntax (e.g., a "syntax of dialogue"), a structural analysis would still not tell us everything that is needed to fully interpret an utterance.

One therefore looks to the third component of an NLP system, semantics. Semantics (or meaning) has typically been formalized in predicate logic, with the assumption that a statement should be interpreted in terms of its "truth value." Logical predicates are precise formulations of the relations between objects in the world (as well as the states of their attributes) and is an elegant approach for representing the canonical meaning of declarative statements which may appear in a variety of different forms. Thus, with the addition of semantics, an NLP system might resolve ambiguities that arise from competing or incomplete phrase structures by referring to a set of predicates that correspond to known facts, or by using logical inferences to derive new facts, in order to discover the truth inherent in a speaker's utterance. Unfortunately, the "truth value" approach is severely challenged by questions, requests, reprimands, and any other statements which are not purely declarative. It requires some stretching of the imagination, for example, to discover the truth being asserted in the expressions, "Excuse me, please" or "Ho, ho, ho."

Logic also fails when an enumeration of predicates cannot represent certain family resemblances implicit in language. For example, one talks about games without being able to list the exact properties that would define one event as a game and another not a game. For this reason, other semantic representations, such as "semantic networks" (Quillian, 1968) and "frames" (Minsky, 1975) have been devised. These approaches can make more arbitrary linkages between objects and attributes than traditional logic would allow, providing "default reasoning" (i.e., guessing) and *ad hoc* inference strategies that deftly ignore logical consistency. But whether it is seen in terms of truth or collections of associations, any interpretation of the "pure" meaning of a given utterance is tainted by the broader context of the speaker-listener interaction. Semantics blends into pragmatics with the recognition that people engaging in conversation actively manipulate each other, qualify what they say, and rely heavily on the history of interaction to phrase subsequent utterances.

While not as well researched as syntax, or even semantics, pragmatics is ultimately the basis from which an NLP system can tell us why a given utterance is emitted, which is of considerable practical importance. However, before examining the pragmatics perspective in general, it may be instructive to first get a feel for how the pragmatics component of an NLP system would typically be used and examine a few specific techniques.

Pragmatics in NLP Systems

Pragmatics has not been the central concern of most NLP systems. Only after ambiguities arise at the syntactic or semantic levels are the context and purpose of the utterance considered for analysis. Consider a problem in which pragmatics has been used in this kind of "support" capacity: ambiguous noun phrases.

Ambiguous Noun Phrases

One of the most common problems in interpreting natural language inputs to a computer system involves the resolution of ambiguous references in noun phrases. For example, if an NLP system encounters the definite reference "the X," there may be no indication in the current sentence as to which X is being discussed. It may be a reference to something that was introduced into the discourse earlier, or this sentence may contain the first mention of X. Without this information, the system may associate improbable or even wildly incompatible attributes to such an object. A similar problem can arise with indefinite noun phrases. A specific indefinite reference, such as "an X" does not identify any particular X and may not have been referenced earlier in the dialogue.

Perhaps even more difficult to interpret is a reference to a non-specific indefinite noun as in "a rose is a rose is a rose." No actual rose is presumably being named here; instead, it is any given instance of the generic class "rose" about which something is being predicated. Finally, and perhaps the most severe problem for the disambiguation of a noun phrase, is the case of pronouns such as "it." This kind of reference is the classic one in need of contextual support for rendering a correct interpretation. Even humans, who are the experts in natural language, are known to ask for clarification of what "it" is referring to.

Some Techniques

An in-depth analysis of how the pragmatics component of an NLP system would support the disambiguation of noun phrases is not our goal here. However, examining a few techniques (based on examples in Gazdar & Mellish, 1989) should give the reader a sense of what kinds of variables are involved.

One approach to the noun phrase problem is to use rules which look for specific inconsistencies or contradictions in a context. Every time a new statement is encountered, the rules are then checked for these properties. For example, assume that the following statements are made to a waiter taking drink orders:

"Water is fine. I'd like *it* with ice."

Within an NLP system this situation might be pattern-matched by the following rule:

If X is "with" Y, then X is not Y.

Given this rule, the system would not be misled into concluding that the *it* in the second sentence refers to ice (e.g., "I'd like ice with ice.").

Another technique is the use of "scripts" —representations of prototypical sequences of events that constrain the possible roles played by actions that occur in a given context. As actions are executed (in this case, as utterances), they are assigned to roles defined in the script; they are said to fill or match the "slots" that represent the events which are expected to flow in a given order. Consider a fragment of a script which has to do with requesting drinks, possibly at a local tavern. Assume that after five stereotypic actions have already occurred, the sixth is that of ordering a drink, represented by Slot #6 below:

Slot #6: (Order a drink): <x>
Slot #7: (<x> with/without ice): [True/False]

The disambiguation of the referent of it in "I'd like it with ice," would presumably be handled by the fact that Slot #7 is "expecting" a reference to whether ice was wanted or not with the beverage ordered in Slot #6. While scripts can have a more general flavor, the example should suggest how many variations of scripts could be proliferated to match a given situation. For highly stereotyped situations, only a few variations may be needed.

A general tool, which would be used by a number of techniques under discussion, is that of "history lists." This involves keeping a detailed list of previous utterances, so that when an ambiguous reference arises (as in the case of the pronoun it) the list is scanned for relevant contextual cues as to its likely referent. One heuristic useful in this vein is to calculate the distance (in words) from a current reference to some previous, unambiguous referent. In the example of ordering water with ice, the noun "water" is a fairly short distance from the pronoun it, and might therefore be determined to be its referent. However, to stretch the point some, the intervening "I'd" might imply that the speaker's intention is to be packed in ice:

"I'd like it (myself) with ice."

History lists can also be used in recognizing the role an ambiguous reference might play in a hierarchically structured task. A task that is hierarchically conceived may have optional sequences of particular elements; as long as all of the task's constituent sub-goals have been satisfied, it may not matter in what order they occur. A simple calculation of distance from an ambiguous pronoun may not find the appropriate referent which is many words removed, but a hierarchical task approach might.

Modeling the Speaker

Beyond the patchwork of specific techniques to handle problems that slip by the syntactic and semantic components of an NLP system, a more thorough-going approach to pragmatics is possible. One might start by constructing a detailed model of the individual speaker in order to interpret the proper force of utterances, whether they are declarative statements, requests, commands, questions, and so forth. Consider the simple question, "Is the apartment too cold?" It could be interpreted as requiring that the listener merely answer "yes" or "no," in which case the speaker may be a landlord asking a new tenant whether the utilities had been turned on yet. But in another context, the question may be asked by a guest in the tenant's apartment who is really asking if the host would please turn up the heat. For either interpretation, we need to understand the speaker in terms of his or her current motivations, the current environmental circumstances, and a significant amount of history.

By considering detailed models of individual members of linguistic communities, we have gone far beyond what has been accomplished in most NLP systems. Certainly, the raw computing power available for modeling whole speakers is a constraint, as well as the known programming techniques which could handle such a task. But neither may help much if the theoretical foundation upon which such a system would be built is inappropriate for the task. Unfortunately, many NLP researchers would agree that "there is still no satisfactory theory of the kinds of pragmatic functions that linguistic utterances can have or quite what is involved in cooperative conversation" (Gazdar & Mellish, 1989, p. 364). The present paper will suggest that the foundation for such a theory already exists, but does not derive from the usual sources that have fueled NLP work.

THE PRAGMATICS PERSPECTIVE

Theories of language can be conveniently characterized in terms of groups of advocates who have emphasized a particular component in what has been described here as a kind of generic NLP system.

Aside from the need to have a division of labor in any complex undertaking, there are definite biases towards syntax, semantics, or pragmatics, with a concommitant subordination of the other components of language. Let us briefly tour the pragmatics perspective, and see what follows.

The philosopher Charles Peirce, who is usually remembered for his insights into logic and semantics (e.g., see Sowa, 1984), coined the term "pragmatics" in the context of linguistic analysis, but it was Austin's (1962) book, *How to do things with words*, based on his lectures at Harvard, that created a definable group of pragmaticist philosophers. Austin's title expressed the force of what language utterances are all about: doing things. Directly building on Austin's work, the philosopher John Searle wrote *Speech Acts* (1969), the book most often cited by pragmaticist philosophers.

For most linguistically-oriented philosophers, however, it is semantics or meaning which occupies the linguistic limelight (e.g., Russell, 1940; but cf. Givon, 1989). Semanticists are primarily concerned with the nature of truth, reference, assertion, and predication; both syntax and pragmatics serve these "higher" goals. A radical pragmaticist would subordinate these goals to understanding the context and purpose of speech acts. Regarding the semanticist's concern for what a sentence is "truly saying," the pragmaticist might answer that sentences do not say anything: people do, and they do so for a reason. After all, do we communicate in order to state truths or do we speak truths (occasionally) so we can communicate?

Not long after Searle's *Speech Acts*, certain linguists revolted against Chomsky's computational/syntactic approach, and became pragmatics advocates (for an analysis of Chomsky's impact on linguistics in America, see Newmeyer, 1986). In particular, developmental linguists, such as Bates (1976), were concerned that efforts to teach language were hurt by Chomsky's nativist program. As Julia (1983) has put it, Chomsky seemed to believe that if a child were surrounded by television sets, s/he would develop language — through a built-in language acquisition device — without ever experiencing the consequences of speaking in the world. Without the proper emphasis on the practical, social, and even political dimensions of language, pragmaticist linguists would claim that there can be no useful guide for educating children.

Of course, for Chomsky (e.g., 1972) syntax is of primary interest, particularly as it provides clues to the cognitive capacities of the human mind/brain. But from a pragmatics point of view, syntax is important only to the extent that it helps achieve some action; "grammaticality" is viewed as a mere refinement. In contrast to what a syntax-oriented linguist might conjecture, grammaticality does not provide any particularly interesting clues as to how the brain is structured; rather, it is more revealing of what works to get speakers and listeners what they want from each other.

While there are huge variations in different pragmaticists' expressed views, a common theme emerges: pragmaticists are working towards a functional analysis of language. Speaking is but one aspect of the behavioral repertoires of some of Earth's organisms and is in principle no different than any other adaptive behavior, such as running or fighting, sexual behavior, foraging, and so forth. It is therefore remarkable that pragmaticists from philosophy and linguistics seem to have so little knowledge or interest in Skinner's *Verbal Behavior* (1957), the most radically pragmatic theory of language to date.

VERBAL BEHAVIOR

Skinner (1957) made a point of dissociating his work from almost anything else which preceded it. Even the apparently pragmatic notion that words are something people use to communicate was to be rejected: people do not "use words" any more than they "use a reach" to grasp something (p. 7). In either case, there is simply behavior, controlled by some combination of historical and current environmental variables. Perhaps Skinner's tendency to never align his work with any other school of thought is what made *Verbal Behavior* such an easy target of Chomsky's (1959) review of the book; in the linguistics community, the review seems to have been more influential than the book. On the other hand, perhaps Skinner did not dissociate himself from earlier works enough, since his critics seem to attack a "behaviorism" that is remarkably un-Skinnerian.

What mades Skinner's contribution especially interesting in the current context is that there has never been a conscious application of the theory of verbal behavior to NLP. This is perhaps because of Skinner's polemic style of writing, as well as the effect of Chomsky's scathing review. In any event, the theory is not known by those in the field of artificial intelligence (cf., West & Travis, 1991). What unique concepts or distinguishing features of the theory of verbal behavior might change the way that NLP systems are conceived, particularly in light of the need for a comprehensive theory of pragmatics?

First, verbal behaviors are classified functionally in terms of their controlling variables. Classes of verbal behavior can be large or small, simple or composed, spoken or written, but they are true functional units only if they are affected by their consequences — reinforcing or punishing, depending on whether the consequence would increase or decrease the likelihood of an instance of that class of behavior occurring again. The purpose of an utterance, a central concern of pragmatics, is therefore an inherent part of classifying all forms, not just the difficult cases. Obviously, the categorization of behavior is not determined by form alone — it is anything but "context free." This clearly relegates both syntactic and fixed semantic structures to a subordinate position; while stock constructions and standard references may be efficiently handled by traditional syntactic and semantic NLP mechanisms, the general rule is that structure and meaning emerge from function.

Second, it should be recognized that verbal behavior is multiply determined; in other words, for each utterance, there is a multitude of historical and current environmental variables (including the speaker's own behavior) that influence its occurrence. A complement to this "convergent" property is that the same environmental variable can momentarily strengthen (make more likely) many different behaviors (i.e., divergently). Thus, there is no one-stimulus, one-response formula. The pragmatics notion of the context of an utterance is richly represented in the theory of verbal behavior and, drawing from the experimental analysis of behavior, even has quantitative aspects that could exploit the computational basis of NLP systems. Multiple control also means that multiple phrase structures and meanings are not considered leftover problems for a separate pragmatics post-processor. The contextually rich nature of all verbal behavior would make pragmatics considerations an on-going concern of the system.

Third, not all of behavior, verbal or otherwise, is observable at the overt level. Events in the environment can momentarily strengthen several different behaviors, only one of which (or some combination) actually surfaces. There may be many behaviors which are only raised to the incipient or covert level, situations we might describe in retrospect by saying, "I almost laughed" or "I said silently to myself." The controlling variables for behavior may not be overt in the environment either; they may involve stimuli felt only within the speaker's body. An implication of this feature is that it may not be enough for effective NLP systems to regard computer interactions as disembodied language elements that occur independently of the people emitting them. A relevant chunk of the person's whole repertoire must be modeled somehow to interpret the otherwise unpredictable forms that arise when both public and private events participate in the production of observable behavior. In fact, while individual sentences and propositions have an apparent life of their own in traditional syntactic and semantic analyses, practical considerations have increased concern for modeling the user in modern NLP systems (Kobsa and Wahlster, 1991).

Combining the second and third features leads us to the existence (in the sophisticated speaker) of autoclitics — responses that qualify, compose, edit, and otherwise make more effective the primary verbal behavior evoked by the environment. Coming in from the hot sun, we might have a tendency to blurt out "Water!" but we can autoclitically compose the response, "May I have some water?" (especially in polite company) and even add "...please" to further ensure the receipt of the water. Skinner (1957, chap. 13) shows that grammar and syntax are derived from such autoclitic processes. Furthermore, Skinner uses the autoclitic to analyze the operations of assertion, quantification, and negation that form the logic of semantics (pp. 322-330). Again, it appears that the pragmatic functions of language are more fundamental than those on which most NLP systems are focused. The evolution of NLP systems seems to be following a progression from focusing on lexical, to syntactic, to semantic, and finally to pragmatic issues probably because the areas attacked earliest were the easiest to implement, not because they were more fundamental. Even the lexical/morphological component of NLP may be reevaluated from the standpoint of minimal verbal operants or response "fragments" that are a function of pragmatic variables.

The dynamics through which verbal behavior is evoked and autoclitically modified are shown in Figure 1. The ordinate is labeled "Strength," which applies to both the controlling variables and the response fragments which they evoke; the abscissa is a crude representation of time. Several elementary verbal relations are strengthened by the presence of these controlling variables, but what will be emitted depends on their combinations; some compete with and weaken each other, others strengthen each other. Autoclitic processes may transform these elementary relations into acceptable speech acts, although autoclitic effects are by no means inevitable. For example, the raw response "Fire!" is generally quite effective as it is, and is probably more effective than "Would you please mind exiting the building which is currently on fire?" But when responses are only strengthened to the incipient level, this may have the effect of evoking certain autoclitic processes, such as composition; thus, a speaker is aware that s/he is about to say something ineffective and refines it to make it more successful.

Stronger than incipient is covert behavior, which is emitted subvocally, allowing us to quickly rehearse before saying something aloud; one can then autoclitically modify it before it reaches the overt level, which is what the listener hears. Of course, one can edit verbal behavior even after it has already become overt, as in the admission, "what I really meant to say was..." Anyone who has ever typed on a computer keyboard can appreciate the enormous amount of overt editing that would be accessible to an NLP system for analyzing more than just the final products of autoclitic processes.

Eight Elementary Relations

Skinner's classification scheme for eight elementary verbal relations, based on typical controlling variables, is depicted in the decision tree in Figure 2. Starting at the top, we see that a motivational variable, such as an aversive situation (e.g., fire) or deprivation (e.g., thirst) will strengthen what would be classified a mand relation (e.g., "Water!"). Many computer commands can be interpreted as mands, although considerable autoclitic activity is required for complex commands. On the opposite branch is verbal behavior controlled by a discriminative stimulus, some part of the environment which is correlated with a higher probability of reinforcement for a given action. This control may be highly abstract, as even the grammaticality of an utterance may function discriminatively (Zuriff, 1976).

This leads us to the next major branch of the tree, between verbal and nonverbal discriminative stimuli, the latter being the basis of the tact (e.g., saying "water" in the presence of a glass of water). Tacts gener-

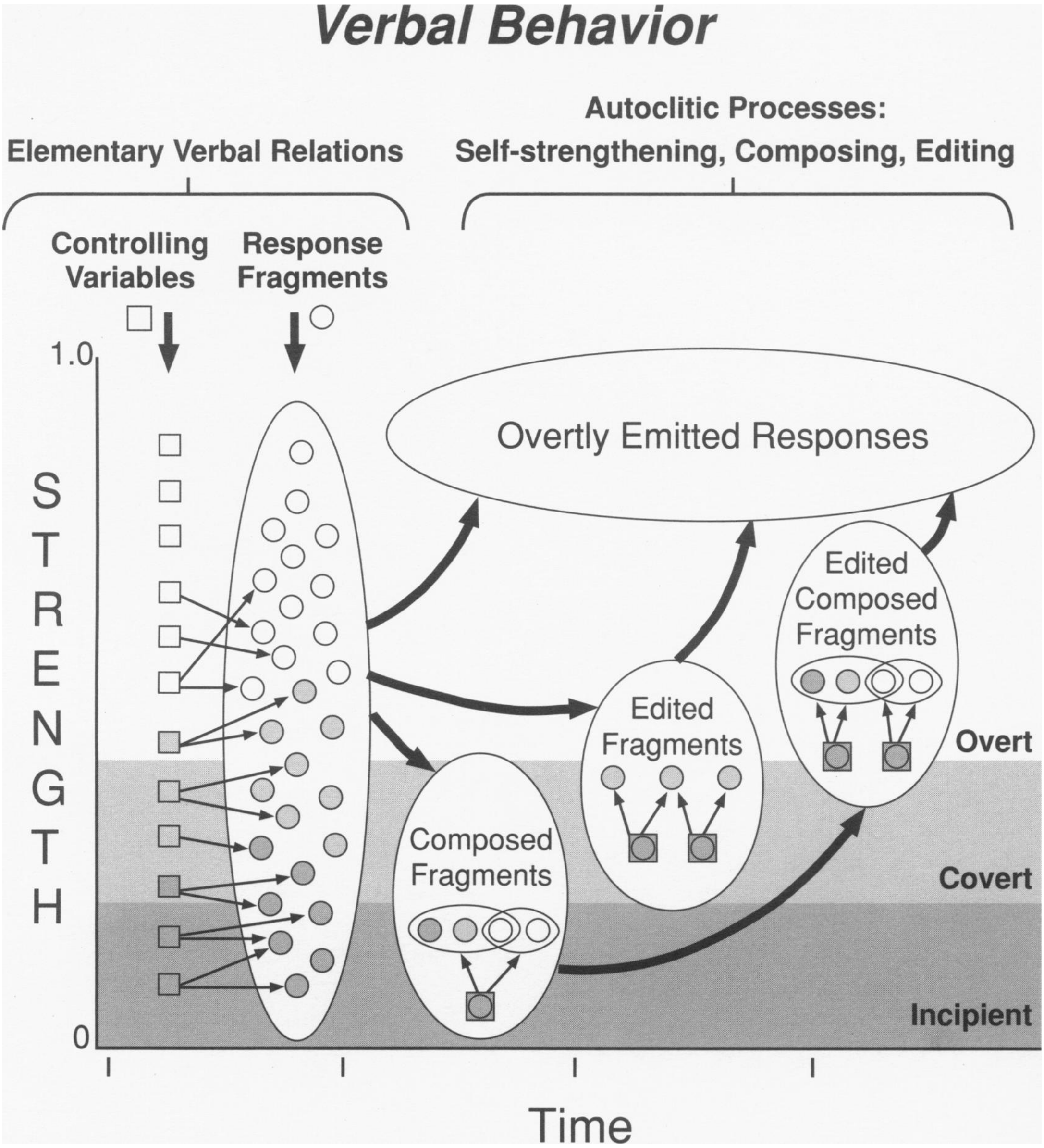

Fig. 1. The dynamics through which verbal behavior is evoked. Response fragments which are strengthened may subsequently serve as controlling variables for composed or edited verbal behavior.

ally benefit listeners more than speakers and have therefore played a smaller role for the human half of human-computer interactions; however, we might metaphorically say that a computer is tacting when it announces to the user that it has just run out of storage space.

The next branch of tree in Figure 2 represents the split between those verbal stimuli which evoke a response having a point-to-point correspondence with those that do not, the latter being labeled an intraverbal (e.g., saying "water" in the presence of someone saying "bread and..."); the association is arbitrary (some might say symbolic) with respect to the structural properties of the verbal stimulus. Computer-based instruction frequently uses intraverbal (thematic) prompts to establish the rich network of intraverbal associations that constitute so much of intellectual repertoires.

The decision tree is completed with the remaining subtypes of verbal behavior

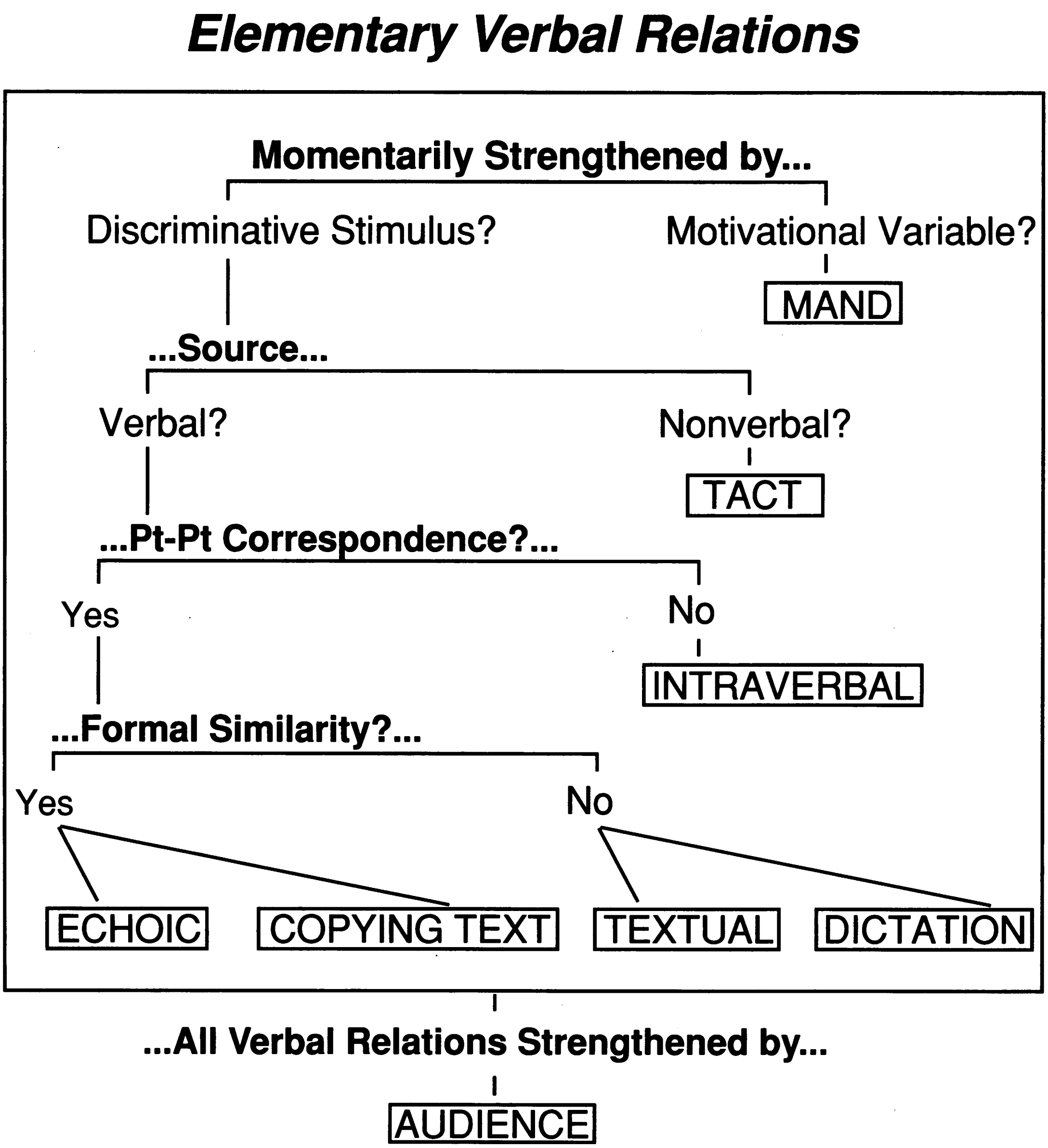

Fig. 2. Decision tree for classifying verbal behavior. All such relations occur in a broader context of audience control.

evoked by verbal stimuli, including those whose stimuli and response products are formally similar. Saying "water" when someone says "water" is echoic, whereas writing "water" in the presence of its written form is copying text. Such formal similarity is lacking in the textual type (saying "water" in the presence of its written form) and in taking dictation (writing "water" when someone says it). One goal of an NLP system would be to reduce the amount of copying and dictation required when using a computer, although one's verbal behavior may still be controlled by these variables even after these relations are no longer required by the input/output limitations of current computer systems.

The eighth relation, a kind of "super" relation, is the audience. The existence of different audiences acts to partition one's verbal repertoire, based primarily on what kind of audience will reinforce or punish one set of responses but not others. One very important phenomenon is that people can act simultaneously as both speaker (or writer) and listener (or reader). This capability is critical to autoclitic processes. The audiences one encounters while working at a computer can be based on the particular language one is programming in (and this "audience" can complain mercilessly if one's syntax is incorrect), the person(s) to whom one is sending an electronic mail message at the moment, or may be more implicit, as when writing an article to submit to a journal.

It would be possible to construct a grammar of verbal behavior in the manner of context-free or tranformational rules. Consider the following sample of rules:

Audience + deprivation —> mand response
Audience + nonverbal stimulus —> tact response

However, one difference from the usual rules found in a classic syntactic analysis is that hybrids of these rules would have to be applied simultaneously because verbal behavior is multiply controlled. In other words, several elementary relations may be involved in any single instance of verbal behavior and no one rule would uniquely recognize such an event. In any case, one would not want to assume that there are actually rules which cause the verbal behavior observed, except for certain circumstances we have not discussed so far.

Rules

A complicating, yet important, factor is the existence of genuine rule-governed behavior. We are often capable of engaging in new behaviors without ever having been exposed previously to the behavioral contingencies which the rules describe. When we are learning to drive a car, we can follow the directions of an instructor in the passenger seat when told, for example, to turn right at the light, even though we have never engaged in such behavior in this situation before. Being directed like this seems to involve the evocation of new combinations or orderings of already familiar behaviors by verbal stimuli, and, in fact, people have usually had extensive histories of being directed in a wide range of tasks by the time they reach driving age.

However, when someone gets instructions on where to turn before ever leaving the house, the situation is somewhat different. The instructions are not part of the current situation in which the driver actually behaves. The rules do not function to direct the driver; they alter his or her repertoire so that a future situation is responded to appropriately. Such rule-governed behavior is defined in *Verbal Behavior* as instruction. But in Skinner's (1969) and others' subsequent writings, the distinction between directions and instructions has been blurred by the term "rule-governed" (see Schlinger, 1990 for a recent review of the issue). The difficulty of separating these two effects is compounded by the fact that we sometimes learn to state a rule before confronting the situation to which it applies and then self-direct ourselves by reciting the rule when in the new situation.

Rules are a mainstay of knowledge representation in artificial intelligence systems, particularly "expert systems" — computer programs which mimic the repertoires of experts in narrow, but important domains, such as diagnosis and scheduling. Experts are interviewed and/or observed while engaged in solving problems by a "knowledge engineer," who then represents this repertoire in a combination of frames and rules. However, practical knowledge engineers do not actually believe that the experts are rule-governed when they solve problems; the rule formulation is just a simplistic way of representing well-differentiated, discriminated operant behavior. Pure cognitive scientists, on the other hand, abound with theories in which virtually all behavior is controlled by (internal) rules.

THE BEHAVIORIST CHALLENGE

Cognitivists generally take a mechanistic approach to language, Chomsky's language acquisition device being one example. Just as a machine can be completely described without recourse to a behaviorist's seemingly endless list of historical and environmental variables, a cognitivist invokes elegant internal representations to connect history with a current situation, presumably in anticipation of neurological findings about the proximal causes of behavior. Computers are well understood machines, and cognitivists are justified in pointing to successes in expert systems and NLP systems as evidence for the validity of using the computer metaphor in emulating intelligent behavior. However, as Schlinger (1992) points out in this journal, these successes have fallen short of the grand visions of AI researchers and recently it appears that the machine metaphor is running a little low on power.

Behaviorists, on the other hand, have some catching up to do. For a group accustomed to pointing to its technological achievements, there is not much behaviorist artificial intelligence that can compare to the successes of expert systems (for notable exceptions see Hutchison & Stephens, 1987; Bell, Hutchison, & Stephens, 1992). A behavioristic NLP system would presumably elevate the NLP system's pragmatics component from its usual afterthought status to a central position, and the theory of verbal behavior should powerfully address the contextual issues that concern pragmaticists.

Proposal: The User Expert

We have seen how solving the problems of pragmatics in NLP systems involves some way of preserving the context of an interaction to disambiguate and guide interpretation. Human factors experimenters do study computer users over a period of hours and intelligent computer-based tutorials may model a student for the duration of an instructional program. But what is missing in all systems is a truly longitudinal perspective.

What is proposed here is a "user expert" — an expert system whose area of expertise is the individual computer user. The behavioral model which the user expert would maintain could conceivably draw from historical interactions which the user has had over the course of an entire career if necessary. As the user expert learns more about its subject, it can increasingly exhibit user-adaptive NLP capabilities. While computer systems today require that the user learn an artificial language, a user expert would reverse the direction of enculturation.

Rather than assuming a role for verbal behavior, for example, as stating truths, we could empirically look for ways that "truth" occurs. Instead of second-guessing the syntactic form that an utterance can take with a general language-understanding mechanism, we could empirically look for ways that verbal behavior is actually structured for an individual. Going beyond traditional models which concentrate on some formal subset of language, such as the word, phrase, sentence, or even "discourse," we can try to understand the person — one who has a unique history and who behaves in a complex environment.

A user expert system would not only make NLP more effective, it could boost the intelligence of other functions. For example, the "personal digital assistants" (PDAs) discussed by Stephens and Hutchison (1992) in this journal, are artificial agents which are dedicated to achieving the goals of the individual computer user. Implicit in this concept is that knowledge of the user determines how the PDA retrieves, presents, or otherwise processes information. Stephens and Hutchison propose that such agents operate according to behavioral principles. A user expert concept by itself does not require that the analysis of the user be based on behavioral principles or that as an intelligent agent it must itself emulate a behavioristic model. The first requirement, however, is the conclusion to which the entire foregoing argument has been headed: the user expert, to be effective, will also be an expert behavior analyst. The second requirement concerns the implementation of the system.

Implementation Techniques

Hutchison & Stephens (1986, May), demonstrated a behavioral approach to artificial intelligence which made use of an adaptive neural network technology. This approach does a superb job of applying behavioral principles to computer systems, and for the last decade has produced practical applications in what typically have been the domains of expert systems. Other neural network pioneers (e.g., Grossberg, 1988) have demonstrated properties of conditioning and stimulus control that parallel the results of laboratory experiments.

A less well known technology which has the potential for applying behavioral principles to computer systems is genetic algorithms (Holland, 1992; Goldberg, 1989) or genetic programming (Koza, 1992). This paradigm involves the simulation of the evolutionary processes of variation (i.e., recombination and mutation) and selection (i.e., fitness-proportional reproduction). Skinner (1981 and elsewhere) has repeatedly made analogies between the selection of behavior by reinforcement during an individual's lifetime and the selection of genetic properties in evolutionary time. By viewing verbal behavior in terms of populations of utterances which recombine and reproduce new variants, based on their effects on audiences, we can exploit the evolutionary metaphor for analyzing the behavior of computer users. While genetic programming offers a way of simulating such evolutionary processes, a detailed formalization of behavioral principles in terms of variation and selection is still being developed (Cherpas, 1992).

While we would assume that verbal behavior is adaptive, we need not make this assumption in too strong a form. Typically, the overgeneralization of the adaptation concept takes the form of optimization. Ultimately, such thinking leads to the rather useless notion that there is only one, grand, optimizing behavior in an organism's repertoire and it applies to all situations (Herrnstein, 1989). However, formalisms other than optimization can be applied to situations where variations are differentially selected, including the evolutionarily stable strategy (ESS) concept of John Maynard Smith (1982). An ESS is an adaptation that is more successful than other available strategies, and remains stable even when it spreads throughout a population. One could view stable features of language in this way — even in the individual speaker. What syntactic and semantic representations take to be essential structure, an evolutionary analysis would view as a dynamic equilibrium (Palmer & Donohoe, 1992).

Behavioristic NLP and Epistemology

If we take the behaviorist view to its radical extreme, then the resolution of the right way to study language (e.g., cognitivist *versus* behaviorist) will ultimately rest in the analysis of scientists who study language. In other words, the behaviorist must admit that no philosophy, no "-ism" (not even behaviorism), is in control of his or her behavior; the truly behaviorist way to study scientific methodology is to analyze the controlling relations that determine the actions of successful scientists.

In fact, the user expert concept was originally conceived as a way of empirically studying the behavior of computer-using scientists, as a replacement for reconstructed scientific methodologies (Cherpas, 1979). If those who are interested in NLP systems were to research their own verbal behavior, particularly the behavior occurring while interacting with their computer systems, a variety of empirically instantiated models might be spawned and analyzed longitudinally. These individual models might then compete or combine with others. The better models should survive and replicate throughout the NLP community, selected for their superior ability to interpret and generate natural and artificial verbal behavior, and perhaps settling into an evolutionarily stable strategy for understanding scientific practice.

REFERENCES

Austin, J. (1962). *How to do things with words.* Cambridge: Harvard University Press.

Bates, E. (1976). *Language and context: The acquisition of pragmatics.* New York: Academic Press.

Bell, T., Hutchison, W., & Stephens, K. (1992). Using adaptive networks for resource allocation in changing environments. In P. Lisgoa (ed.), *Current perspectives in neural networks.* London: Chapman and Hall.

Cherpas, C. (1979). Beyond behaviorism, *Experimental Analysis of Behavior Colloquium Series.* Kalamazoo, MI.

Cherpas, C. (1992). The heuristic nature of the concepts of variation and selection. Unpublished manuscript.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). Review of B. F. Skinner's *Verbal Behavior. Language, 35,* 26-58.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge: MIT Press.

Chomsky, N. (1972). *Language and mind*. New York: Harcourt, Brace, Jovanovich.

Gazdar, G., & Mellish, C. (1989). *Natural language processing in LISP*. Workingham: Academic Press.

Givon, T. (1989). *Mind, code, and context: Essays in pragmatics*. Hillsdale: Lawrence Erlbaum Associates.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading: Addison-Wesley.

Grossberg, S. (1988). *Neural networks and natural intelligence*. Cambridge: MIT Press.

Herrnstein, R. (1989). Darwinism and behaviourism: Parallels and interactions. *Evolution and its influence*. In A. Grafen (ed.). Oxford: Clarendon Press.

Holland, J. (1992). *Adaptation in natural and artificial systems*. Cambridge: MIT Press.

Hutchison, W., & Stephens, K. (1986, May). Artificial intelligence: demonstration of an operant approach. Paper presented at the annual meeting of the Association for Behavior Analysis, Milwaukee, WI.

Hutchison, W., & Stephens, K. (1987). The airline marketing tactician (AMT): A commercial application of adaptive networking. *Proceedings of the First International Conference on Neural Networks*, Vol 4.

Julia, P. (1983). *Explanatory models in linguistics*. Princeton: Princeton University Press.

Kobsa, A., & Wahlster (1989). *User models in dialog systems*. New York: Springer-Verlag.

Koza, J. (1992). *Genetic programming*. Cambridge: MIT Press.

Minsky, M. (1975). A framework for representing knowledge, In P. Winston (ed.), *The psychology of computer vision*. New York: McGraw-Hill.

Newmeyer, F. (1986). *The politics of linguistics*. Chicago: University of Chicago Press.

Palmer, D., & Donohoe, J. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *The American Psychologist, 47*, 1344-1358.

Quillian, R. (1968). Semantic memory, In M. Minsky (ed.), *Semantic information processing*. Cambridge: MIT Press.

Russell, B. (1940). *Inquiry into meaning and truth*. New York: W. N. Norton.

Schlinger, H. D. (1990). A reply to behavior analysts writing about rules and rule-governed behavior, *The Analysis of Verbal Behavior, 8*, 77-82.

Schlinger, H. D. (1992). Intelligence: Real or artificial? *The Analysis of Verbal Behavior, 10*, 125-133.

Searle, J. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1969). *Contingencies of reinforcement*. Englewood Cliffs, NJ: Prentice-Hall.

Skinner, B. F. (1981). Selection by consequences. *Science, 213*, 501-504.

Smith, M. J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Sowa, J. (1984). *Conceptual structures: Information processing in mind and machine*. Reading: Addision-Wesley.

Stephens, K., & Hutchison, W. (1992). Behavioral personal digital assistants: The seventh generation of computing, *The Analysis of Verbal Behavior, 10*, 149-156.

West, D., & Travis, L. (1991). The computational metaphor and artificial intelligence: A reflective examination of a theoretical falsehood. *AI Magazine, 12*, Spring.

Winograd, T. (1982). *Language as a cognitive process volume I: Syntax*. Reading: Addison-Wesley.

Woods, W. (1970). Transition network grammars for natural language analysis. *Communications of the ACM, 13*, No. 10.

Woods, W. (1973). An experimental parsing system for transition networks grammars. In J. Rustin (ed.), *Natural language processing*. New York: Algorithmics Press.

Zuriff, G. (1976). Stimulus equivalence, grammar, and internal structures. *Behaviorism, 4*, 43-52.